DOCUMENT RESUME

ED 238 710 SE 043 686

AUTHOR Wasserman, Edward A.; Shaklee, Harriet

TITLE Judging Response-Outcome Relations: The Role of Response-Outcome Contingency, Outcome Probability, and Method of Information Presentation.

INSTITUTION Iowa Univ., Iowa City. Dept. of Psychology.

SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE [83]
CONTRACT NIE-G-80-0091
NOTE 59p.; Appendix E of SE 043 682.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cognitive Processes; Educational Research; *Evaluative Thinking; Higher Education; *Learning; Mathematical Concepts; *Probability; Problem Solving; *Psychological Studies

ABSTRACT

Four experiments investigated college students' judgments of inter-event contingency. Subjects were asked to judge the effect of a discrete response (tapping a wire) on the occurrence of a brief outcome (a radio's buzzing). Pairings of the possible event-state combinations were presented in a summary table, an unbroken time line, or a broken time line format. Subjects judged the extent to which the response caused the outcome or prevented it from occurring. Across all methods of information presentation, judgments were a positive function of response-outcome contingency and outcome probability. In the unbroken time line condition, judgments of negative response-outcome contingencies were less extreme than judgments of equivalent positive contingencies. Judgments of positive and negative relationships were generally symmetrical in the summary table condition. Summary table judgments were less influenced by the overall probability of outcome occurrence. These judgment differences among format conditions suggest that, depending on the method of information presentation, subjects differently partition event sequences into discrete event pairings. (Author/MNS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED238710

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Judging Response-Outcome Relations: The Role of Response-Outcome Contingency, Outcome Probability, and Method of Information Presentation

Edward A. Wasserman and Harriet Shaklee

The University of Iowa

Running head: Judging Response-Outcome Relations

SE 043 686

Abstract

A series of four experiments investigated college students' judgments of interevent contingency. Subjects were asked to judge the effect of a discrete response (tapping a wire) on the occurrence of a brief outcome (a radio's buzzing). Pairings of the possible event-state combinations (response-outcome, response-no outcome, no response-outcome, no response-no outcome) were presented in a summary table (Experiments 2 and 4), in an unbroken time line (Experiments 1, 2, and 4), or in a broken time line format (Experiment 3). Subjects judged the extent to which the response caused the outcome or prevented it from occurring. Across all methods of information presentation, judgments were a positive function of response-outcome contingency and outcome probability. In the unbroken time line condition, judgments of negative response-outcome contingencies were less extreme than judgments of equivalent positive contingencies. This asymmetry was smaller in the broken time line condition and in those conditions where subjects were encouraged to segment an unbroken time line into discrete response-outcome units. Finally, judgments of positive and negative relationships were generally symmetrical in the summary table condition. Relative to the two time line portrayals, summary table judgments were also less influenced by the overall probability of outcome occurrence. These judgment differences among format conditions suggest that, depending on the method of information presentation, subjects differently partition event sequences into discrete event pairings. The segmenting of continuous event streams may be an important factor in the accuracy of everyday judgments of interevent contingency.

> And now remains,
> That we find out the cause of this effect,
> Or rather say the cause of this defect,
> For this effect defective comes by cause.
>
> W. Shakespeare; Hamlet, II, ii

Students of behavior both before and after Shakespeare have been interested in causal perception. Most noteworthy was D. Hume (1739) who proposed a set of conditions which were conducive to cause-effect impressions. Hume's insights into the psychology of causation have helped to shape the direction of subsequent research and theory in the area.

Also important have been discussions of causal perception from comparative and developmental perspectives. C. L. Morgan (1893, 1894) concluded on the basis of extremely limited evidence that human adults, but not children and animals, can perceive the relationship between events. More systematic data led Inhelder and Piaget (1958) to propose a stagewise unfolding of the human's conception of interevent correlation or contingency as the individual develops from child to adult.

Subsequent investigations into the perception of interevent relations have not yielded evidence that is consistently favorable to the developmental and evolutionary speculations of Morgan and of Inhelder and Piaget. Nor is the evidence particularly supportive of modern theories, which posit a virtual identity between humans' and animals' perceptions and the actual interevent contingencies that prevail in their environments (e.g., Heider, 1958; Kelley, 1967; Mackintosh, 1974; Rescorla, 1978).

In the basic human judgment paradigm, subjects are given information about the frequency of pairings of alternative states (e.g., presence and absence) of two events (e.g., plant food and plant health); they can then be asked to judge the direction and magnitude of the relationship between the events. In many of these experiments, adults do not accurately judge the correlation between two binary variables (see Crocker, 1981 for a review).

Despite these negative results, other work has been more successful in showing that adults can accurately judge interevent relations under some circumstances (e.g., Allan & Jenkins, 1980; Alloy & Abramson, 1979; Seggie, 1975; Seggie & Endersby, 1972; Shaklee & Tucker, 1980). Nevertheless, many factors have been suggested over the past 20 years which may contribute to distortions in the perception of correlation.

Investigators have found that the accuracy of correlational judgments depends on the sign of the relationship being judged. In particular, Erlick and Mills (1967) found that subjects judged negative correlations as closer to zero than positive correlations of equal magnitudes. Also common is the result that subjects find contingencies of zero to be especially difficult to identify. For example, Seggie (1975) reported that subjects were accurate in their judgments of contingent relationships, but were error-prone in judging noncontingent relationships (also see Allan, 1980; Allan & Jenkins, 1980). Alloy and Abramson (1979) replicated this pattern of differential accuracy in nondepressed subjects, but found that depressed adults judged noncontingent problems closer to zero than did nondepressed subjects.

One must, however, be cautious in interpreting the effects of relationship direction; subjects may approach the stimuli in question with strong expectations about the nature of the relationship that will hold. In Seggie's 1975 study, for example, subjects judged whether or not hospitalizing a victim of a tropical disease would improve the chances of recovery. Erlick and Mills' (1967) subjects judged the relationship between the quantity of a particular food a person ate and whether the person felt better or worse. People who believe in the merits of medical science or hearty eating would be likely to expect each to improve general well being. This expectation could produce a bias to report relationships as positive, resulting in errors in judging negatively related and independent events.

Evidence of such an expectation effect was found in the research of Chapman and Chapman (1967 a & b), where subjects judged there to be a positive relationship between semantically-associated clinical signs and symptoms in stimuli that actually presented the sign and symptom as independent, or even negatively related. This illusory correlation effect proved to be highly resistant to a variety of attempts to reduce it, including exposing subjects to the stimuli several times and offering them a $20 reward for accuracy. Similar expectancy effects may be a reason for some past findings of differential accuracy as a function of relationship direction. Any attempt to examine the effect of relationship direction should then be conducted in a context in which prior expectations are minimal.

A second common finding in past research is that judgments of interevent correlations are biased by the relative frequencies of the event states of the variables involved. For example, Jenkins and Ward (1965) asked subjects how much control their responses (pushing Button 1 or 2) had over the frequency with which a score light appeared. Subjects' judgments of control were most strongly correlated with the number of times the score light occurred, regardless of whether that outcome was actually influenced by their choice of buttons. Allan and Jenkins (1980) found that this bias was reduced, but not eliminated when subjects had a single button to press or not to press, compared to Jenkins and Ward's two-button condition (also see Alloy & Abramson, 1979). The findings of these investigations indicate that the probability of the outcome is a second possible confound to be controlled or manipulated in assessing contingency judgment.

A final recurrent finding in past research is that the accuracy of judging interevent contingency depends on how the event frequency information is presented. Two common formats present this information either as a series of

individual event-state combinations (e.g., Alloy & Abramson, 1979; Shaklee & Mims, 1982; Ward & Jenkins, 1965) or as a summary table (e.g., Seggie, 1975; Smedslund, 1963; Ward & Jenkins, 1965). Experiments which have compared the two presentation formats have found accuracy to be higher when the frequency information is summarized in table format.

Of course, the serial and summary formats differ in a variety of ways. Most obvious is the added memory demand involved in the trial-by-trial presentation of information; thus, subjects who add a strong memory load to an already complex judgment process may compromise accuracy to simplify an overwhelming task. Shaklee and Mims (1982) relied upon such a memory account in interpreting their judgment findings. Ward and Jenkins (1965), however, argued that, while important, memory load cannot fully account for the judgment difference between serial and summary formats. Rather, they proposed that the serial presentation of stimulus information may lead subjects to organize the information differently from those who view the same information in a tabled format. In support of this point, Ward and Jenkins note that subjects in their experiments who were shown tabled information after serial presentation used less appropriate judgment strategies than those who saw only the tabled information. If information is organized differently under the two conditions, then this may lead subjects to make different judgments of inter-event relationships. Although this reasoning is plausible, past paradigms have confounded presentation format with memory load; the contributions of memory and organization effects in past research cannot then be separated. The issue is best addressed by comparing use of serial and summary frequency information in conditions alike in memory load.

The present study thus compared serial and summary formats in a setting free of memory demands, while also using a problem for which subjects should

have little bias as to the nature of the interevent relation. The basic situation involved troubleshooting a malfunctioning radio. While this situation is far less dramatic than Polonius' efforts to determine the reason for Hamlet's odd behavior, it is nonetheless representative of everyday instances of causal reasoning.

Subjects were told that an individual was trying to find the cause of an intermittent buzz (B) by occasionally tapping (T) on a wire inside the radio. The results of the troubleshooting were then given to the subject, who was asked to judge the degree to which tapping affected the radio's buzzing: from "causes the sound to occur" to "has no effect on the sound" to "prevents the sound from occurring." This context has the virtue of being one in which subjects should not have a strong expectation about the nature of the response-outcome relationship; tapping a wire should be as likely to complete as to break a loose connection. Similarly, if the wire is not loose, tapping it should have no effect on the buzz.

Holding constant the probability of tapping, $p(T)$, both the probability of a buzz given a tap, $p(B/T)$, and the probability of a buzz given no tap, $p(B/\overline{T})$, were systematically varied to yield 24 different troubleshooting conditions. These conditions in turn constituted nine tap-buzz contingencies, $p(B/T) - p(B/\overline{T})$, ranging in .25-steps from -1.00 to +1.00 (see Allan, 1980 for further discussion of various measures of contingency or correlation).

An additional feature of the 24 troubleshooting conditions was that they were contrived in such a way that they varied not only in the tap-buzz contingency, but also in the overall probability per sampling interval of the buzzing sound, $p(B)$. Eight different buzz probabilities were studied, ranging in .125-steps from .125 to 1.000. Because the tap-buzz contingency and the relative frequency of the radio's buzzing *vs* its not buzzing were independent

dimensions in the present experimental design, the contributions of these variables to subjects' judgments of correlation could be individually assessed.

The method of information presentation was studied with two basic techniques. In one, subjects were given summary tables showing the numbers of times that the four possible event sequences occurred in 24 sampling intervals: tap-buzz, tap-no buzz, no tap-buzz, and no tap-no buzz. In the other, the same information was given in a time line format, with the 24 sampling intervals graphically and linearly arrayed. Such an arrangement preserves the sequential character of the critical events, while minimizing the strong memory demands that are ordinarily placed on subjects when they are given information in a trial-by-trial fashion. This method was originally suggested by Ward and Jenkins (1965, p. 240); however, it has never been utilized in experimental research.

Since past work has not entailed a time line presentation of event frequencies, our series of investigations began by looking at subjects' judgments using this format alone. Experiment 1 explored the effects of tap-buzz contingency and buzz probability on judgments of tap-buzz correlation in both within-subjects and between-subjects paradigms. Experiment 2 directly compared the effects of the time line and summary table methods of information presentation. Because the second experiment disclosed that judgments did differ under the two conditions of information presentation, Experiments 3 and 4 explored possible reasons for the judgment differences.

Experiment 1

The first experiment investigated the judgment of response-outcome correlation when responses and outcomes were shown to subjects in a time line format. In one part of the experiment, each subject received only 1 of 24 possible tap-buzz conditions; in the other part, each subject received all 24

tap-buzz conditions. Both between- and within-subjects conditions were included in order to identify possible influences of multiple judgments, since we hoped to use the more efficient within-subjects procedure in later work. Subjects' ratings of the response-outcome relationships allowed us to determine the degree to which the tap-buzz contingency, $p(B/T) - p(B/\overline{T})$, and the overall probability of the buzzing sound, $p(B)$, influenced their behavior. To determine whether the sign of the response-outcome correlation affected subjects' judgments, equal numbers of positive and negative contingencies were studied.

Method

Subjects. The subjects were participants in an introductory psychology class, who served in the experiment as one option for fulfilling a course requirement. A total of 552 students served in the between-subjects part of the experiment and a total of 25 students served in the within-subjects part.

Problems. A set of 24 problems was constructed. These problems were alike in that they all comprised 24 sampling intervals. Each sampling interval in turn had two components: a "response" component during which a tap might or might not occur, and an "outcome" component during which a buzz might or might not occur. Each of the 48 resulting components of a problem was denoted on the subject's problem sheet as a dash; the 48 consecutive dashes thus constituted the time line for each problem. Taps in the response component of a sampling interval were denoted by an "A" above the dashed time line, and buzzes in the outcome component of a sampling interval were denoted by a "B" below the dashed time line.

For all 24 problems, there were 12 taps represented in the 24 possible response components. Thus, the probability of tapping per sampling interval, $p(T)$, was always .50. Problems varied in terms of the likelihood that a buzz was represented in the outcome components, $p(B)$, and the likelihood of buzzes following taps, $p(B/T)$, and no taps, $p(B/\overline{T})$, in the response components.

For each of the 24 problems, Table 1 shows the numbers of sampling intervals of each of four possible types: tap-buzz, tap-no buzz, no tap-buzz, and no tap-no buzz. Note that the number of sampling intervals with a tap is equal to 12, which is the same as the number of sampling intervals without a tap. Note also that the total number of sampling intervals equals 24. And note finally that the number of sampling intervals with a buzz varies from 3 to 24.

Insert Table 1 about here

For each problem, time lines were constructed from smaller groupings that contained eight sampling intervals. The sequence of event pairings was determined randomly within each eight-sample group. While eight-sampling groups theoretically provide all the necessary information that is needed to distinguish the 24 problems, we thought it advantageous to triple the amount of input given to the subjects in hopes that their judgments might thereby be improved. For example, Problem 18 in Table 1 was represented as follows:

A	A	A	A	A	A	A			A			A		A	A	A
	B	B		B		B			B	B	B	B	B	B		B

Figure 1 shows a second method of depicting the 24 problems that were studied. Both the top and bottom portions of the figure locate each problem within the unit square defined by the two independent conditional probabilities, $p(B/T)$ and $p(B/\bar{T})$. The top portion of the figure shows the response-outcome contingency, $p(B/T) - p(B/\bar{T})$, of each of the problems; the bottom portion shows the likelihood of the buzzing sound per sampling interval, $p(B)$, for the same problem set. There are nine response-outcome contingencies and eight probabilities of buzz presentation represented by the 24 problems in Figure 1. Furthermore, these two procedural dimensions are orthogonal, as can be seen by

the opposite slopes of the lines that connect the 24 problems in the top and bottom portions of the figure. From the figure it can finally be seen that one possible problem was not included in the set. When $p(B/T) = 0 = p(B/\bar{T})$, $p(B) = 0$; little sense could thus have been made of the task by the subjects (see next section for questionnaire instructions).

Insert Figure 1 about here

Procedure. Subjects were given problem sheets that each contained instructions, a time line, and a rating scale. The instructions read as follows:

> After buying a new radio, Kim finds that it emits a brief buzzing sound every so often. Kim finds this buzzing sound annoying and decides to find its cause. Removing the back of the radio, Kim suspects that a wire may be loose. Kim chooses a wire and taps on it a number of times in order to see if this has any effect on the buzzing sound. In the diagram below, Kim's tapping on the wire is shown by an A above the time line which moves from left-to-right across the page. An occurrence of the brief buzzing sound is shown by a B below the time line.

One of the 24 different time lines then followed. Below the time line was a nine-point rating scale ranging from -4 (prevents sound from occurring) to 0 (has no effect) to +4 (causes sound to occur). Subjects were asked to circle the number that best corresponded to their answer to the question, "If you were Kim, what would you conclude was the effect of tapping on the wire?"

In the between-subjects part of the experiment, only 1 of the 24 problem sheets was given to each subject. In the within-subjects part of the experiment, each subject received all 24 problem sheets, with the order of the sheets randomly determined for each subject. The 24 problem sheets were clipped together; each packet also included the following cover sheet:

> The aim of this experiment is to see how people judge the relationship between their actions and the consequences of those actions. In the 24 sheets that follow, the same basic problem is posed: What is the relation between Kim's tapping on the wire of a malfunctioning radio and the occurrence of a brief buzzing sound that the radio occasionally emits. The 24 sheets differ only in the particular relationship between Kim's tapping and the occurrence of the sound. For each of the 24 sheets, please rate the degree to which Kim's tapping affects the rate of the radio's buzzing, from "prevents the sound from occurring" to "causes the sound to occur." As you go through the 24 problems, you'll soon see that the problems differ from one another to varying degrees. You may sometimes want to look back to prior problems; you may even want to change prior responses. This is OK. It is more important to work through the problems carefully and methodically than to give quick and offhand reactions. Indeed, the materials are paper-clipped together so that you can sort through the many sheets and organize them any way you wish.

Results

Table 2 shows the means and standard deviations of subjects' judgments for the 24 problems in both the between- and within-subjects parts of the experiment. Each of the 24 problems is located in the table by the coordinates $p(B/T) - p(B/\overline{T})$ and $p(B)$. In general, subjects' rating scores were positive functions of both $p(B/T) - p(B/\overline{T})$ and $p(B)$.

Insert Table 2 about here

Figure 2 graphically portrays subjects' rating scores as separate functions of $p(B/T) - p(B/\overline{T})$ and $p(B)$ in each part of the experiment. Analysis of variance simultaneously assessed the reliability of these two sets of functions.

Insert Figure 2 about here

The left panel of Figure 2 displays subjects' ratings as a function of $p(B/T) - p(B/\overline{T})$. The positive diagonal in the figure shows the responses of a

hypothetical judge whose responses correspond in a linear fashion to the actual response-outcome contingencies and who also employs the full rating scale. In the between- and within-subjects parts of the experiment, subjects' judgments were reliable linear functions of $p(B/T) - p(B/\overline{T})$, $F(1, 528) = 139.17$, $p < .001$, and $F(1, 24) = 74.76$, $p < .001$, respectively; however, the slopes of those functions were clearly less than that of our hypothetical linear observer. The between- and within-subjects functions also had reliable quadratic components, $F(1, 528) = 11.28$, $p = .001$, and $F(1, 24) = 28.07$, $p < .001$, respectively; this trend appears to be due to the negative segments of the functions having shallower slopes than the positive segments. Finally, in the within-subjects part of the experiment, the contingency-rating function had a reliable cubic component, $F(1, 24) = 10.96$, $p = .003$; this trend appears to be due to the function having an inverted S shape. Although the overall form of the between-subjects function was similar, it did not have a reliable cubic component.

The right panel of Figure 2 displays subjects' ratings as a function of $p(B)$. In the within-subjects part of the experiment, ratings were a positive linear function of $p(B)$, $F(1, 24) = 32.63$, $p < .001$. In the between-subjects part of the experiment, the linear trend only approached significance, $F(1, 528) = 2.90$, $p = .089$.

To assess the relative contributions of $p(B/T) - p(B/\overline{T})$ and $p(B)$ to subjects' judgment scores, the percentage of problem variance accounted for by these factors was determined through the cubic component of each; beyond the cubic component, no significant variance remained for either part of the experiment. In the between-subjects part of the experiment, $p(B/T) - p(B/\overline{T})$ accounted for 86.47% of the total variance and $p(B)$ accounted for 3.21%; in the within-subjects part of the experiment, the corresponding scores were 71.87% and 24.10%.

Discussion

Subjects' judgments of contingencies in the time line format showed several interesting trends that were generally comparable in the within- and between-subjects parts of the experiment. These results also accord well with past paradigms using different presentation formats. First, judgments of response-outcome correlation were a reliable function of the contingency between the tapping of a wire and the occurrence of a brief buzzing sound. Subjects' ratings rose as the tap-buzz contingency, $p(B/T) - p(B/\overline{T})$, increased from negative to positive values. Thus, subjects clearly showed some sophistication about appropriate bases of contingency judgment.

The relative accuracy of subjects' judgments is, however, another issue. Mean judgments indicated that subjects rated noncontingent relationships close to zero, but ratings of several negative relationships hovered close to zero as well. While subjects were asked to rate both the degree and the sign of a correlation, the clearest evidence of accuracy here was the rated direction of the relationship. Subjects' judgments should also have been ordered according to the strength of the correlation. While this was generally true, the ratings yielded contingency judgments that were poorer than ideal. Indeed, the quadratic component of the judgment function indicates that subjects did not treat positive and negative relationships symmetrically; contingencies of the same absolute value were rated as stronger for positive than for negative relationships. The form of this difference in ratings of relationship strength closely resembles that found in prior research by Erlick and Mills (1967).

The second main finding was that judgments of correlation were reliably influenced by the likelihood of the buzzing sound, $p(B)$. This bias is comparable to that found in other studies in which the judgment of contingency depended on the likelihood that the outcome occurred (Allan & Jenkins, 1980;

Alloy & Abramson, 1979; Jenkins & Ward, 1965). These prior studies most convincingly demonstrated a bias effect of p(B) with response-outcome contingencies of zero; Allan and Jenkins' (1980) investigation further suggested that the bias effect could arise under positive contingencies. The present report confirms the above trends and also shows that the effect of p(B) on judgments holds under negative response-outcome contingencies as well (see that ratings tend to increase from top to bottom within most columns of Table 2).

Experiment 2

The results of the time line portrayals in Experiment 1 were comparable in many ways to those of past paradigms. However, subjects who view information in a particular format may treat the information in a manner specific to that format; that is, subjects' attention to information may depend on the way the information is presented. The organization or integration of attended information may vary with stimulus format as well. We propose three ways in which the time line and the more familiar summary table format may produce different judgments.

First, tabled presentation of event frequency information offers the subjects tallies of the frequencies of each type of event-state combination. Our time line presentation (like past serial presentation techniques) requires the subjects to generate such tallies on their own. Subjects given time line information may guess rather than count those frequencies, resulting in estimation errors. This logic suggests that judgments with time line presentation will be generally less accurate than judgments with tabled presentation and that such differential accuracy will be relatively constant across positive, negative, and noncontingent relationships. The resultant judgment function should be relatively flat across all contingencies compared to that of tabled information.

A second possible source of difference is the fact that the summary table presents the event-state combinations in a form of comparable salience. In contrast, each type of event pairing has a unique representation in the time line format (i.e., AB, A-, -B, --). As a result, some types of event pairings may be more salient than others. In particular, the interval pairs with two event absences (--) may be less prominent than those with one or both events present. This feature may also have been true of past serial presentation paradigms. If so, subjects should underestimate the frequency of no tap-no buzz pairings. Since the denominator of the conditional probability, $p(B/\overline{T})$, would then be smaller than would be accurate, this would result in an estimate of $p(B/\overline{T})$ that is too high. This in turn should result in a bias to judge contingencies as being more negative in the time line format than the same contingencies presented in the tabled format.

Finally, the time line format allows the subject to determine the delay between tap and buzz that will be counted as a tap-buzz pairing. Consider the interval series A--B. The tabled format would represent this as one occurrence of tap-no buzz and one of no tap-buzz. However, a subject given the time line presentation may well consider this series to be a single pairing of tap-buzz. This tendency would lead to an underestimation of the frequencies of event pairings tap-no buzz and no tap-buzz and an overestimation of the frequency of tap-buzz pairings. These errors would yield an inflated numerator for $p(B/T)$ and a smaller than accurate numerator for $p(B/\overline{T})$. These biases should result in judgments of contingencies being more positive in the time line than in the summary table format. This problem of event segmenting should not have been true of past discrete trial presentations, where each slide or card defined an event-outcome pairing. However, the problem may be true of event processing in real time, when event continua must be defined as discrete events.

Thus, each of three reasons for judgment differences in the two information presentation conditions would result in a unique pattern of judgment outcomes. Whether any of these differences will materialize is an empirical question. Experiment 2 addressed this issue by comparing judgments under the time line format employed in Experiment 1 with judgments of the same problems presented in the summary table format used in past investigations (e.g., Smedslund, 1963; Ward & Jenkins, 1965). Since judgments were so comparable in the between- and within-subjects parts of Experiment 1, subjects in Experiment 2 judged all 24 problems.

Method

Subjects. The subjects were 34 undergraduate research participants.

Problems. The same 24 problems were used here as in Experiment 1. Problems in the time line format were typed on a single sheet of paper with the nine-point rating scale to the right of each problem. Problems in the summary table format were typed on another sheet of paper similar to Table 1, except that the four types of sampling intervals were vertically arrayed; identical rating scales were located beneath each problem. Problems were presented in a single random sequence for the time line format and in a different random sequence for the table format.

Procedure. During the first portion of the experimental session, subjects were given an instruction sheet describing the troubleshooting problems on the attached sheet of paper. For half of the subjects the problems were in the time line format, and for the other half the problems were in the summary table format. During the second half of the session, subjects worked problems in the format not worked in the first half. Instructions for time line problems were the same as those used in Experiment 1. Instructions for summary table problems were the same, with appropriate adjustments to introduce the table rather than the time line format.

Results

Table 3 shows the means and standard deviations of subjects' judgments for the 24 problems given in the time line and summary table formats. Because analysis of variance failed to disclose any reliable effects attributable to the order of format presentation, this factor is not considered in Table 3 nor in later data analysis. As in Experiment 1, subjects' ratings were positive functions of both $p(B/T) - p(B/\overline{T})$ and $p(B)$.

Insert Table 3 about here

Figure 3 graphically depicts subjects' rating scores as separate functions of $p(B/T) - p(B/\overline{T})$ and $p(B)$ for each method of information presentation. Analysis of variance simultaneously compared these two sets of functions.

Insert Figure 3 about here

The left panel of Figure 3 portrays subjects' ratings as a function of $p(B/T) - p(B/\overline{T})$. Overall, ratings were reliable linear, $F(1, 32) = 51.72$, $p < .001$, and quadratic, $F(1, 32) = 12.90$, $p = .001$, functions of tap-buzz contingency. Additionally, there was a reliable quadratic contingency by format interaction, $F(1, 32) = 4.97$, $p = .033$. To pinpoint the source of this interaction, separate analyses of variance were conducted on the time line and summary table data. For both the time line and the summary table formats, ratings were reliable linear functions of contingency, $F(1, 33) = 36.77$, $p < .001$, and $F(1, 33) = 44.27$, $p < .001$, respectively. However, the quadratic trend was reliable for the time line format only, $F(1, 33) = 14.59$, $p = .001$. Thus, subjects' judgments were reliable linear functions of response-

outcome contingency with both methods of information presentation; however, the method of information presentation influenced those functions, with the tabled format supporting judgments that better approximated those of an ideal observer, particularly in the region of negative contingencies.

The right panel of Figure 3 illustrates subjects' ratings as a function of $p(B)$. Overall, ratings were reliable linear, $F(1, 32) = 30.11$, $p < .001$, and quadratic, $F(1, 32) = 26.68$, $p < .001$, functions of outcome probability. Additionally, there were reliable linear, $F(1, 32) = 6.32$, $p = .017$, and quadratic, $F(1, 32) = 12.99$, $p < .001$, outcome probability by format interactions. Because of these interactions, follow-up analyses were separately performed on the time line and summary table data. For the time line data, ratings were reliable linear, $F(1, 33) = 34.57$, $p < .001$, and quadratic, $F(1, 33) = 30.43$, $p < .001$, functions of $p(B)$; for the summary table data, the linear trend was reliable, $F(1, 33) = 5.33$, $p = .027$, and the quadratic trend fell just short of statistical significance, $F(1, 33) = 3.69$, $p = .063$. Thus, the method of information presentation altered the influence of outcome probability on subjects' ratings; providing the information in a time line format both steepened the probability-judgment function and increased its curvature relative to providing the same information in a summary table format.

And, regardless of tap-buzz contingency and buzz probability, judgments were reliably higher in the time line condition than in the summary table condition, $F(1, 32) = 5.03$, $p = .032$.

To assess the relative contributions of response-outcome contingency and outcome probability to subjects' ratings, the percentage of problem variance accounted for by each factor was determined as in Experiment 1. For the summary table data, $p(B/T) - p(B/\overline{T})$ accounted for 81.35% of the total variance and $p(B)$ accounted for 12.58%; for the time line data, the corresponding

scores were 39.48% and 51.79%. Beyond the cubic component, no significant variance remained for the summary table data. For the time line data, the 8.78% remaining variance was small, but statistically significant, $F(17, 561) = 3.23$, $p < .001$.

Discussion

The data from subjects given the time line in this experiment replicate the judgment patterns of subjects in the comparable condition of Experiment 1. In addition, the results of Experiment 2 confirm prior findings (Shaklee & Mims, 1982; Smedslund, 1963; Ward & Jenkins, 1965) that the method of information presentation affects subjects' judgments of response-outcome correlation.

The obtained judgment differences under two conditions comparable in memory demands suggest that past effects of presentation conditions may not be solely attributed to memory. In general, subjects' judgments were more closely attuned to response-outcome contingency when information was given in the summary table than when the same information was given in the time line. First, the contingency-judgment function (left panel of Figure 3) was more symmetrical about zero in the summary table condition, suggesting that subjects rated positive and negative relationships in a comparable fashion. Again, the time line portrayal supported less accurate judgments of negative than positive contingencies. Second, table format judgments were less distorted by the probability of the buzzing sound (right panel of Figure 3). The linear contingency by format interaction showed that the time line judgments were steeper functions of $p(B)$ than the summary table judgments.

We previously reviewed three reasons why time line and summary table formats may result in different contingency judgments. The suggestion that the time line will lead to more errors in estimating frequencies of event

pairings than the summary table predicted overall poorer contingency judgment accuracy (i.e., a flatter, but symmetrical contingency-judgment function) in the time line than in the tabled format condition. The possibility that joint event absences (no tap-no buzz) were less salient in the time line than in the tabled presentation mode predicted a general bias to report relationships as more negative in the time line than in the summary table format. However, neither of these difference patterns describe our results.

Subjects in this experiment did show a tendency to judge relationships as more positive in the time line than in the summary table condition. This result supports our third proposed source of differences, that subjects may group event pairings differently in the time line than the tabled format. In particular, event series A--B could be identified as a single tap-buzz occurrence rather than a tap-no buzz and a no tap-buzz, yielding in a bias to report relationships as positive. However, we should note that while ratings were generally higher in the time line than in the summary table condition, the positivity bias was more pronounced for negative than positive contingencies. One possible account for this finding involves the influence of context on the grouping of event pairings; that is, A--B may be most likely to be judged a tap-buzz occurrence when there are few contiguous AB pairings in the time line, as would be the case in negative contingencies.

Besides helping us to understand why different presentation formats support different judgments, these performance differences between groups also allow us to reject the possibility that time line subjects' problems with rating negative contingencies are due to a response bias or to prior expectations. Any expectation about the effect of tapping on the radio's buzzing should be the same in the two groups, but judgments of negative contingencies were distorted for time line subjects only. Similarly, since subjects made

judgments on the same rating scale in the two conditions, performance differences cannot be attributed to peculiarities in the scale itself.

Experiment 3

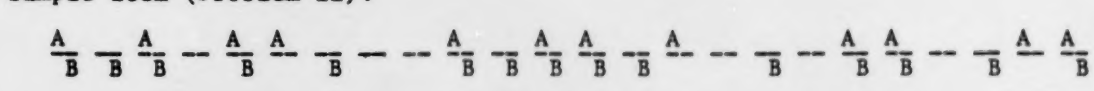
The results thus far suggest that subjects may define events differently in the time line and table formats. If this is the principal reason for the inaccurate responses of time line subjects, then their judgments should improve when the continuous stream of events in the time line is separated into discrete units.

Our third experiment further explored the problem of defining individual sampling periods by placing a clear break between paired intervals in the time line format. To do this, we simply added a blank space between successive sampling intervals along the time line. As in the within-subjects part of Experiment 1, subjects rated all 24 tap-buzz contingencies. These judgments were compared to those obtained in Experiment 1, in which successive sampling intervals immediately followed one another.

Method

Subjects. Another group of 25 undergraduate research participants joined the 25 who had served in the within-subjects part of Experiment 1, and whose data are depicted again in the Results section that follows. Subjects in these two groups were from the same introductory psychology course and were tested within 3 weeks of the same school term.

Problems. The problems for the new subjects were identical to those in Experiment 1, except that one blank space was inserted between successive sampling intervals along the time line. This format is illustrated in a sample item (Problem 11):



Procedure. The procedure for the new subjects given the broken time lines was identical to that for the former subjects given the unbroken time lines in Experiment 1.

Results

Table 4 shows the means and standard deviations of subjects' judgments for the 24 problems given in the broken and the unbroken time line conditions of Experiment 3. Again, subjects' ratings were positive functions of $p(B/T) - p(B/\overline{T})$ and $p(B)$.

Insert Table 4 about here

Figure 4 graphically illustrates subjects' rating scores as separate functions of $p(B/T) - p(B/\overline{T})$ and $p(B)$ for each time line condition. Analysis of variance simultaneously compared these two sets of functions.

Insert Figure 4 about here

The left panel of Figure 4 shows subjects' ratings as a function of $p(B/T) - p(B/\overline{T})$. Overall, ratings were reliable linear, $F(1, 576) = 542.75$, $p < .001$, quadratic, $F(1, 576) = 34.32$, $p < .001$, and cubic, $F(1, 576) = 20.35$, $p < .001$, functions of tap-buzz contingency. Additionally, there was a reliable linear contingency by time line interaction, $F(1, 576) = 5.08$, $p = .025$, and a near significant quadratic contingency by time line interaction, $F(1, 576) = 3.18$, $p = .075$. Therefore, separate analyses of variance were conducted on the data for the group given the broken time line and for the group given the unbroken time line. For both the broken and unbroken time line groups, ratings were reliable linear functions, $F(1, 24) = 83.74$, $p < .001$,

and $F(1, 24) = 74.76$, $p < .001$, respectively; quadratic functions, $F(1, 24) = 7.17$, $p = .013$, and $F(1, 24) = 28.07$, $p < .001$, respectively; and cubic functions of contingency, $F(1, 24) = 24.83$, $p < .001$, and $F(1, 24) = 10.96$, $p = .003$, respectively. Thus, although the contingency-rating functions were similar, judgments of contingency were more strongly differentiated for subjects in the broken time line group; this greater differentiation was generally more notable for negative than for positive contingencies.

The right panel of Figure 4 portrays subjects' ratings as a function of $p(B)$. Overall, ratings were reliable linear, $F(1, 576) = 139.87$, $p < .001$, and quadratic, $F(1, 576) = 25.33$, $p < .001$, functions of outcome probability. Additionally, there was a reliable quadratic outcome probability by time line interaction, $F(1, 576) = 6.18$, $p = .013$. Separate analyses of variance were therefore conducted on the data from the two time line groups. For both the group given the broken time line and the group given the unbroken time line, ratings were reliable linear functions of $p(B)$, $F(1, 24) = 20.62$, $p < .001$, and $F(1, 24) = 32.63$, $p < .001$, respectively. However, the quadratic trend was reliable for the broken time line group only, $F(1, 24) = 24.01$, $p < .001$. Thus, the probability-rating functions of the two time line groups were similarly sloped, although the function for the broken time line appeared to turn downward at high outcome probabilities more than the function for the unbroken time line.

To assess the relative contributions of response-outcome contingency and outcome probability to subjects' judgments, the percentage of variance accounted for by each factor was determined as in Experiments 1 and 2. For the broken time line group, $p(B/T) - p(B/\overline{T})$ accounted for 77.31% of the total problem variance and $p(B)$ accounted for 19.08%; for the unbroken time line group, the corresponding scores were 71.37% and 24.10%.

Discussion

We introduced the broken time line format in Experiment 3 to partition the time line continuum into discrete sampling intervals. The results of the experiment indicate that this manipulation had an effect on judgments of the problem set. Subjects judging broken time lines showed greater differentiation in their ratings as a function of the scheduled contingency than subjects judging unbroken time lines. This increased differentiation was generally more prominent for negative than for positive relationships, a difference which was also true of subjects judging tabled information in Experiment 2.

Thus, the results of subjects who viewed the broken time lines duplicate in some respects the behavior of subjects judging on the basis of tabled information. Our ability to increase the accuracy of contingency judgments by this manipulation enhances confidence in our interpretation that subjects made errors in identifying discrete event pairings in the continuous time lines. The similarity of judgments of tabled and broken time line information suggests that one function of the table may be to separate a stream of events into coherent units. Such units may be more readily classed according to the type of event pairing and thus may be more accurately incorporated into a contingency judgment.

While breaking the flow of the time line into discrete sampling intervals yielded judgments more similar to those made with summary table presentation, inspection of Figures 3 and 4 shows that the judgments obtained under these two conditions were not identical. Contingency-judgment functions under the broken time line format were less symmetrical about zero than under the summary table format, and probability-rating functions were steeper in the former condition than in the latter. Thus, other factors may well contribute to the differences in contingency judgments obtained with the time line and summary table formats in Experiment 2.

Experiment 4

Thus far, our leading interpretation of the problems created by a continuous representation of events is that people have difficulty breaking the stream into discrete units. An alternative approach to testing this account might be to teach people to parse the time line into the component units. If such training produces judgment functions like those found in our broken time line and table formats, such findings would further support this as the source of judgment differences. A second function of the table mentioned earlier might be to offer subjects numerical summaries of the information about the four event combinations. This summarized information may be more readily incorporated into a decision rule in judging event covariations. In this way, judgment accuracy might be further enhanced if subjects were asked to count the occurrences of each event-state combination and note these frequencies in a table. By this process, subjects would effectively convert a time line into a table format.

Our fourth and final experiment used each of these approaches. One group of subjects was presented with the 24 problems in our original time line format, but were taught to break the line into response-outcome intervals (line-interval). A second group received these instructions and were also asked to count the frequencies of each event-state pairing and write those frequencies in a table (line-table). Time line and table groups using our original instructions served as comparison conditions for these manipulations. Improved judgment by line-interval subjects compared to time line subjects would further implicate line segmenting as a factor in contingency judgment. Further improvements by line-table subjects would suggest that summary information is also an important function of the tabled format. Because we found sex differences in contingency judgment in related work of ours (Shaklee & Hall, in press), sex was included as a factor in this experiment.

Method

Subjects. A total of 160 introductory psychology subjects served in the experiment with 20 males and 20 females in each of four judgment conditions.

Problems. The 24 contingencies for this experiment were the same as those in the previous experiments. Format of problems in the time line and table representations was the same as that used in Experiments 1 and 2.

Procedure. The introduction to the troubleshooting problems was identical to that used in the previous studies, except that the problem representation was explained in one of four ways:

Line: These instructions were the same as those used in Experiments 1 and 2.

Line-Interval: These problems were represented in a time line like that used in Experments 1 and 2, but in this case subjects were specifically instructed how to break the time line into response-outcome intervals. Instructions were as follows:

> Each dash on the time line represents one unit of time. Time units come in pairs, with the first an opportunity for a response (Tap or No Tap) and the second an opportunity for an outcome (Buzz or No Buzz). Thus, pairs of successive intervals can be of four types: Tap-Buzz, Tap-No Buzz, No Tap-Buzz, No Tap-No Buzz. For each of the time lines, please rate the degree to which Kim's tapping affects the rate of the radio's buzzing, from "prevents the sound from occurring" to "causes the sound to occur."

Line-Table: Problems and instructions were identical to those in the Line-Interval condition, except that each problem was accompanied by a blank table labeled as in the previous table condition of Experiment 2. Subjects were instructed to complete the table before making their judgment. Instructions were as follows:

> Each dash on the time line represents one unit of time. Time units come in pairs, with the first an opportunity for a response (Tap or No Tap) and the second an opportunity for

an outcome (Buzz or No Buzz). Thus, pairs of successive intervals can be of four types: Tap-Buzz, Tap-No Buzz, No Tap-Buzz, No Tap-No Buzz. For each time line, please count the frequency of each of these four types of interval pairs. Enter those frequencies in the table to the right of the time line. Once you have completed the table, please rate the degree to which Kim's tapping affects the rate of the radio's buzzing, from "prevents the sound from occurring" to "causes the sound to occur."

Table: Problems and instructions in this condition were identical to those in Experiment 2.

In each condition, the information offered in the instructions was shown on a sample problem illustrating each type of response-outcome pairing. Subjects were invited to ask any questions they might have, after which they proceeded at their own pace through the problem set.

Results

Means and standard deviations of subjects' judgments for the 24 problems in each judgment condition are shown in Table 5. Figure 5 illustrates subjects' judgments of the nine contingencies, $p(B/T) - p(B/\overline{T})$, and the eight probabilities of buzzing sound, $p(B)$, for the four judgment conditions. These functions were simultaneously compared by analysis of variance, including sex of subject and judgment condition as factors. Paired follow-up analyses were conducted on interactions, setting alpha at .025 to reduce the experiment-wide error rate.

Insert Table 5 and Figure 5 about here

The overall analysis yielded reliable linear, $F(1, 152) = 851.86$, $p < .001$, quadratic $F(1, 152) = 100.92$, $p < .001$, and cubic $F(1, 152) = 12.52$, $p < .001$ trends of response-outcome contingency on subjects' judgments. As in our previous experiments, judgments were a function of problem contingency, but

with judgments of negative relations closer to zero than those of positive relations. This analysis also showed a main effect of judgment condition, $F(3, 152) = 11.40$, $p < .001$, although that effect is qualified by a contingency by condition interaction, $F(23, 3496) = 2.47$, $p < .001$. As seen in the left portion of Figure 5, the form of this interaction shows that judgments in the Table condition were most symmetrical about zero, judgments in the Line condition were least symmetrical, and judgments in the Line-Interval and Line-Table conditions fell between these two extremes. Follow-up analyses compared contingency judgment functions for selected condition pairs. Line-Interval and Line conditions were compared to identify the effect of the interval segmenting instructions. This analysis showed Line-Interval subjects to be significantly different from Line subjects; linear trend $F(1, 76) = 11.12$, $p = .001$, the quadratic trend approaching significance $F(1, 76) = 4.92$, $p = .029$. Comparison of Line-Table and Line-Interval contingency functions showed that tabling the frequency information had no additional effect on judgment accuracy. Line-Table and Table judges were compared to see if judges who tabled the frequency information for themselves were equivalent in judgment to those who judged tables provided by the experimenter. This comparison showed that contingency judgment functions were not equivalent for the two groups, with Line-Table and Table judges reliably different in quadratic trend, $F(1, 76) = 5.83$, $p = .018$, but not in linear or cubic trends.

Sex differences in contingency functions were statistically significant, with the contingency-judgment function for females flatter than that for males: linear trend $F(1, 152) = 3.94$, $p = .049$, cubic trend $F(1, 152) = 4.38$, $p = .038$. This sex effect did not interact significantly with judgment condition.

As in our previous experiments, subjects' judgments were an increasing function of the probability of the buzzing sound (see right portion of Figure

5). Ratings showed significant linear, $F(1, 152) = 210.66$, $p < .001$, quadratic, $F(1, 152) = 80.90$, $p < .001$, and cubic, $F(1, 152) = 4.58$, $p = .034$, trends as a function of $p(B)$. Unlike previous analyses, however, these probability-judgment functions were not reliably affected by judgment condition, although the Line group again showed the greatest effect of $p(B)$ and the Table group showed the least effect. Effects of $p(B)$ also did not differ as a function of subjects' sex.

The relative contributions of response-outcome contingency and outcome probability in each of the four conditions were determined as in the prior experiments. For the Table group, $p(B/T) - p(B/\overline{T})$ accounted for 89.07% of the total problem variance and $p(B)$ accounted for 9.47%; for the Line-Table group, the corresponding scores were 80.97% and 17.02%; for the Line-Interval group, the scores were 76.04% and 17.61%; and for the Line group, the scores were 71.38% and 22.64%. In only the latter two groups was the residual variance significant: Line-Interval residual = 6.35%, $F(17, 646) = 6.72$, $p < .001$, and Line residual = 5.98%, $F(17, 646) = 2.25$, $p = .003$.

Since frequency judgment errors may detract from contingency judgment accuracy, the frequency tables generated by subjects in the Line-Table condition were examined for accuracy. Overall, errors were small, with mean absolute deviations of .15, .10, .30, and 1.65 for Tap-Buzz, Tap-No Buzz, No Tap-Buzz, and No Tap-No Buzz frequencies, respectively. In view of the differential judgments of positive and negative relationships in this condition, frequency judgment accuracy was compared for problems representing positive and negative contingencies. Absolute deviations were averaged across table cells for this analysis. A matched-pairs t-test showed no reliable differences in frequency judgment errors on positive and negative contingencies, $t(39) < 1$.

Discussion

Experiment 4 represents a conceptual replication of our third experiment.

In Experiment 3, we broke the time line into discrete units. In this experiment, we taught the subjects themselves to define these intervals. The results indicate that the manipulations in the two experiments had similar effects. Line-Interval and Line-Table subjects in Experiment 4 produced contingency-judgment functions intermediate to those of our Line and Table subjects. Line-Interval and Line-Table subjects' contingency-judgment functions were more symmetrical about zero than that of Line subjects, although the two new conditions did not differ from each other. This failure to find additional improvement by subjects who completed a frequency table indicates that the availability of summary information contributes little to judgment accuracy. However, the similarity of these two functions to that of subjects in our past broken time line condition enhances our confidence in the problem of event segmenting as a source of error in judging negative relationships.

The finding that Line-Table judges are also less accurate than Table judges is a bit of a surprise. These subjects have effectively converted time line information into a tabled format. However, the accuracy of that conversion is a second question. Since any deviations in frequency judgments must necessarily be in the direction of lower accuracy, subjects in this condition may have somewhat erroneous information on which to base their judgments. However, a look at subjects' frequency counts indicates reasonable accuracy; indeed, 12 out of 40 subjects did not show a single error on any of the 24 problems. In addition, error rates were similar on negative and positive contingency problems. Thus, inaccuracy of frequency judgments constitutes a weak account of the difference in judgment functions of Line-Table and Table subjects.

These differences between Line-Table and Table judgments replicate the stimulus presentation effects of Ward and Jenkins (1965) in a substantially

different format. Their subjects viewed sequences of event-outcome pairs (cloud seeding or not/rain or no rain), each sequence indicating some degree of positive relationship. When the sequence was complete, one group of subjects saw a table summarizing the frequencies of each of the event-state combinations. A second group saw the tabled information only. Ward and Jenkins found that subjects who saw the tabled information after the event series were less accurate in their judgments than those who saw the tabled information alone. It was this finding that inspired the experimenters to conclude that viewing the event sequence had caused the subjects to represent the information in a way that the table failed to counteract, perhaps differentially emphasizing the relative importance of particular event-state pairings. Our own results parallel these past findings closely. In our case, however, subjects viewed event contingencies in a linear representation free of memory demands.

As in our previous experiments, subjects' judgments here were biased by the probability of the buzzing sound. However, unlike Experiments 2 and 3, the extent of that bias was not reliably different in the Line and Table judgment conditions. The failure to replicate this finding is surprising and difficult to account for given the comparability of other aspects of the present results to our other previous outcomes. This finding does temper our confidence in the previous result that judgments of tabled information are relatively free of the effect of the probability of outcome.

Finally, this experiment showed a reliable effect of sex, with contingency-judgment functions of females reliably flatter than those of males. This difference may indicate that females have a higher judgment error rate than males, contributing to flatter functions. This interpretation is congruent with findings in our related work (Shaklee & Hall, in press) showing

that females use simpler, less accurate rules than those used by males to judge event covariations. An alternative interpretation of the sex differences in the present experiment is that the two sexes judge the problems with similar accuracy, but that the females use a more limited range of the scale to make their judgments. However, a comparison of judgments indicates that the two sexes use the scale extremes ($\pm$4) at comparable rates (11.3% and 12.2% of judgments for males and females, respectively), ruling out response conservatism as a viable account of this sex difference.

Concluding Comments

In overview, the results of four different experiments suggest that judgments of interevent contingency importantly depend on the method of presenting information about event pairings. Most accurate were judgments of summary table information (Experiments 2 and 4); least accurate were judgments of information presented in a continuous time line format (Experiments 1, 2, and 4). The accuracy of subjects judging partitioned time lines (Experiment 3) fell in between that of the other two conditions. Subjects trained to segment continuous time lines (Experiment 4) made judgments similar to those who saw partitioned time lines. This evidence suggests that Ward and Jenkins (1965) were correct in their suspicion that presentation format may influence subjects' treatment of frequency information in making contingency judgments. Our evidence indicates that subjects may break event sequences into different discrete event pairings depending on the format in which the frequency information is presented. This explanation accounts well for our own findings, but may not be similarly useful in explaining the effects of relationship direction in some past paradigms. As noted earlier, slide or card sequence presentations offer event pairings as discrete units rather than as event continua.

This interpretation offers a ready account for the finding in past research that subjects judge negative relationships less accurately than

positive relationships. Past researchers have suggested that subjects know how to judge positive, but not negative contingencies. Allan (1980), however, pointed out one difficulty with this interpretation; subjects who only know how to judge positive relationships must be able to distinguish between positive and negative relationships in order to apply the appropriate rule to positive contingencies. Presumably, a different, less accurate rule is applied to negative (and independent) relationships. Thus, this interpretation requires that an individual maintain more than one rule to judge event contingencies, and that the person know when to apply which rule to which relationship.

Our analysis indicates a single judgment problem which would result in differential accuracy on positive and negative relationships: that is, subjects' boundaries for event segments depend on the other events in the stream. Positive relationships are typified by many response-outcome pairs which would define a brief time interval as a response-outcome unit. However, where few outcomes promptly follow responses, the observer may accept relatively delayed outcomes as "caused" by the response. The estimate of response-outcome pairs is inflated, resulting in an illusion of a relationship which is less negative than is objectively the case.

We would argue that the problems our subjects encountered in the time line format could be similar to those encountered in judgments of real world contingencies--response-outcome delays may vary in everyday experience. One task of the perceiver is then to define which sequences represent true response-outcome pairings. Investigations of the cues used to break event sequences into discrete units are rare. Our evidence suggests that understanding this process may be important to our ability to account for contingency judgments.

References

Allan, L. A note on measurements of contingency between two binary variables in judgment tasks. Bulletin of the Psychonomic Society. 1980, 415, 147-149.

Allan, L., & Jenkins, H. The judgment of contingency and the nature of the response. Canadian Journal of Psychology/Review of Canadian Psychology, 1980, 34, 1-11.

Alloy, L., & Abramson, L. Judgment of contingency in depressed and nondepressed students: Sadder but wiser? Journal of Experimental Psychology: General, 1979, 108, 441-485.

Chapman, L., & Chapman, J. Genesis of popular but erroneous psychodiagnostic observations. Journal of Abnormal Psychology, 1967, 72, 193-204. (a)

Chapman, L., & Chapman, J. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. Journal of Abnormal Psychology, 1967, 74, 271-280. (b)

Crocker, J. Judgment of covariation by social perceivers. Psychological Bulletin, 1981, 90, 272-292.

Erlick, D. E., & Mills, R. G. Perceptual quantification of conditional dependency. Journal of Experimental Psychology, 1967, 43, 9-14.

Heider, F. The psychology of interpersonal relations. New York: Wiley, 1958.

Hume, D. A treatise of human nature. In A. Flew (Ed.), On human nature and the understanding. New York: Collier, 1962. (A Treatise of human nature was originally published in 1739.)

Inhelder, B., & Piaget, J. The growth of logical thinking from childhood to adolescence. New York: Basic Books, 1958.

Jenkins, H., & Ward, W. Judgment of contingency between responses and outcomes. Psychological Monographs, 1965, 79, 1-17.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), Nebraska Symposium on Motivation (vol. 15). Lincoln: University of Nebraska Press, 1967.

Mackintosh, N. J. The psychology of animal learning. London: Academic Press, 1974.

Morgan, C. L. The limits of animal intelligence. Fortnightly Review, 1893. 54, 223-239.

Morgan, C. L. An introduction to comparative psychology. London: Walter Scott, Ltd., 1894.

Rescorla, R. A. Some inplications of a cognitive perspective on Pavlovian conditioning. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), Cognitive processes in animal behavior. Hillsdale, NJ: Erlbaum, 1978.

Seggie, J. The empirical observation of the Piagetian concept of correlation. Canadian Journal of Psychology/Review of Canadian Psychology, 1975, 29, 32-42.

Seggie, J., & Endersby, H. The empirical implications of Piaget's concept of correlation. Australian Journal of Psychology, 1972, 24, 3-8.

Shaklee, H., & Hall, L. Methods of assessing strategies for judging covariation between events. Journal of Educational Psychology, in press.

Shaklee, H., & Mims, M. Sources of error in judging event covariations: Effects of memory demands. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1982, 8, 208-224.

Shaklee, H., & Tucker, D. A rule analysis of judgments of covariation between events. Memory and Cognition, 1980, 8, 459-467.

Smedslund, J. The concept of correlation in adults. Scandinavian Journal of Psychology, 1963, 4, 165-173.

Ward, W., & Jenkins, H. The display of information and the judgment of contingency. Canadian Journal of Psychology, 1965, 19, 231-241.

Footnote

This research was supported by NSF grant 79-14160 to E.A.W. and NIE grant G-80-0091 to H.S. The authors are greatly indebted to R. H. Hohle for his helpful technical assistance. Portions of this research were reported at the annual meeting of the Psychonomic Society, Philadelphia, PA, November, 1981. Requests for reprints may be sent to either author, Department of Psychology, The University of Iowa, Iowa City, IA 52242.

Table 1

Frequencies of Response-Outcome Possibilities
in Each Experimental Problem

Problem	Tap-Buzz	Tap-No Buzz	No Tap-Buzz	No Tap-No Buzz
1	12	0	0	12
2	9	3	0	12
3	6	6	0	12
4	3	9	0	12
5	12	0	3	9
6	9	3	3	9
7	6	6	3	9
8	3	9	3	9
9	0	12	3	9
10	12	0	6	6
11	9	3	6	6
12	6	6	6	6
13	3	9	6	6
14	0	12	6	6
15	12	0	9	3
16	9	3	9	3
17	6	6	9	3
18	3	9	9	3
19	0	12	9	3
20	12	0	12	0
21	9	3	12	0
22	6	6	12	0
23	3	9	12	0
24	0	12	12	0

Table 2

Means and Standard Deviations (in Parentheses) of Subjects' Ratings in the Between- and Within-Subjects Parts of Experiment 1

	$p(B/T) - p(B/\bar{T})$								
p(B)	-1.00	-0.75	-0.50	-0.25	0.00	+0.25	+0.50	+0.75	+1.00
Between Subjects									
.125				-1.57 (1.53)		0.13 (0.90)			
.250			-0.91 (1.59)		0.09 (1.84)		1.30 (1.57)		
.375		-1.04 (2.07)		-0.74 (1.48)		0.17 (1.79)		1.61 (1.52)	
.500	-1.43 (2.10)		0.00 (1.87)		-0.13 (2.05)		0.96 (1.49)		2.30 (2.37)
.625		-0.39 (1.81)		-0.52 (2.00)		0.39 (1.69)		1.78 (1.69)	
.750			-0.30 (1.97)		0.00 (1.14)		1.61 (1.44)		
.875				-0.52 (2.02)		0.09 (2.02)			
1.000					0.09 (0.88)				
Within Subjects									
.125				-1.48 (1.36)		-0.52 (1.65)			
.250			-0.60 (1.94)		-0.60 (1.72)		0.88 (1.63)		
.375		-0.92 (1.85)		-0.48 (1.10)		0.40 (0.94)		1.96 (1.31)	
.500	-1.16 (1.78)		0.00 (1.36)		0.08 (1.49)		1.52 (1.45)		3.48 (1.42)
.625		0.20 (1.39)		0.12 (1.27)		1.28 (1.22)		2.24 (1.24)	
.750			0.44 (1.39)		0.60 (1.20)		2.12 (1.11)		
.875				1.28 (1.46)		1.48 (1.58)			
1.000					0.92 (1.90)				

Table 3

Means and Standard Deviations (in Parentheses) of Subjects' Ratings Under the Time Line and Summary Table Formats of Experiment 2

	p(B/T) - p(B/$\bar{T}$)								
p(B)	-1.00	-0.75	-0.50	-0.25	0.00	+0.25	+0.50	+0.75	+1.00
Time Line									
.125				-2.38 (2.06)		-2.09 (2.11)			
.250			-1.09 (2.05)		-1.15 (2.20)		0.56 (2.19)		
.375		-1.32 (1.81)		-0.62 (1.78)		0.94 (1.24)		1.41 (1.97)	
.500	-0.94 (2.11)		-0.26 (1.38)		-0.06 (1.75)		1.29 (1.74)		2.47 (2.29)
.625		0.62 (1.85)		0.32 (1.34)		1.29 (1.15)		1.85 (1.80)	
.750			0.71 (1.72)		0.85 (1.54)		1.85 (1.77)		
.875				1.76 (2.04)		1.62 (2.00)			
1.000					0.79 (2.26)				
Summary Table									
.125				-1.41 (2.18)		-0.21 (2.18)			
.250			-1.09 (2.72)		-0.38 (1.91)		0.74 (2.36)		
.375		-1.03 (2.55)		-1.03 (2.02)		0.29 (1.49)		1.26 (1.82)	
.500	-1.44 (2.17)		-0.74 (1.87)		0.24 (1.06)		1.15 (1.65)		2.44 (1.87)
.625		-1.68 (1.74)		-0.06 (1.55)		1.03 (1.54)		1.24 (2.38)	
.750			-0.29 (2.01)		0.50 (1.54)		1.62 (1.97)		
.875				0.38 (2.26)		0.91 (2.12)			
1.000					0.50 (1.74)				

Table 4

Means and Standard Deviations (in Parentheses) of Subjects' Ratings Under the Broken and Unbroken Time Line Conditions of Experiment 3

	$p(B/T) - p(B/\bar{T})$								
p(B)	-1.00	-0.75	-0.50	-0.25	0.00	+0.25	+0.50	+0.75	+1.00
Broken Time Line									
.125				-1.64 (1.98)		-0.48 (1.98)			
.250			-1.36 (1.62)		-0.28 (1.37)		0.96 (1.56)		
.375		-0.96 (2.09)		0.36 (1.16)		0.56 (1.17)		2.24 (1.77)	
.500	-2.12 (2.63)		0.16 (0.83)		0.36 (1.32)		1.60 (1.36)		3.80 (0.98)
.625		-0.60 (1.96)		0.52 (1.02)		1.68 (1.26)		2.08 (1.57)	
.750			0.12 (1.39)		1.12 (1.34)		1.92 (1.44)		
.875				0.84 (1.41)		1.52 (1.47)			
1.000					0.24 (0.86)				
Unbroken Time Line									
.125				-1.48 (1.36)		-0.52 (1.65)			
.250			-0.60 (1.94)		-0.60 (1.72)		0.88 (1.63)		
.375		-0.92 (1.85)		-0.48 (1.10)		0.40 (0.94)		1.96 (1.31)	
.500	-1.16 (1.78)		0.00 (1.36)		0.08 (1.49)		1.52 (1.45)		3.48 (1.42)
.625		0.20 (1.39)		0.12 (1.27)		1.28 (1.22)		2.24 (1.24)	
.750			0.44 (1.39)		0.60 (1.20)		2.12 (1.11)		
.875				1.28 (1.46)		1.48 (1.58)			
1.000					0.92 (1.90)				

Table 5

Means and Standard Deviations (in Parentheses) of Subjects' Ratings in the Four Conditions of Experiment 4

	$p(B/T) - p(B/\bar{T})$								
p(B)	-1.00	-0.75	-0.50	-0.25	0.00	+0.25	+0.50	+0.75	+1.00
Line									
.125				-1.93 (1.97)		-0.78 (2.27)			
.250			-0.78 (1.84)		-0.55 (1.72)		1.15 (1.77)		
.375		-0.98 (1.93)		-0.45 (1.73)		0.70 (1.99)		2.13 (1.60)	
.500	-1.28 (1.95)		-0.25 (1.32)		0.05 (1.53)		1.58 (1.53)		3.45 (1.16)
.625		0.45 (1.84)		0.25 (1.32)		1.23 (1.33)		2.25 (1.32)	
.750			0.55 (1.99)		0.55 (1.72)		1.60 (1.77)		
.875				0.60 (2.30)		1.83 (1.66)			
1.000					0.68 (2.08)				
Line-Interval									
.125				-2.33 (1.52)		-0.33 (2.04)			
.250			-2.10 (1.69)		-0.58 (1.46)		1.48 (1.38)		
.375		-1.80 (1.93)		-0.60 (1.26)		1.28 (0.89)		2.60 (1.02)	
.500	-2.55 (1.48)		-0.80 (1.31)		0.63 (1.07)		1.70 (0.95)		3.80 (0.64)
.625		-0.15 (1.73)		0.23 (1.15)		1.15 (1.11)		2.70 (0.81)	
.750			0.63 (1.20)		0.63 (1.09)		2.13 (1.05)		
.875				0.80 (1.68)		2.08 (1.49)			
1.000					0.20 (1.42)				

Table 5 (continued)

p(B)	p(B/T) - p(B/$\bar{T}$)								
	-1.00	-0.75	-0.50	-0.25	0.00	+0.25	+0.50	+0.75	+1.00
Line-Table									
.125				-1.90 (1.39)		-0.70 (1.91)			
.250			-1.60 (2.31)		-0.63 (1.43)		0.58 (1.50)		
.375		-2.48 (1.60)		-1.30 (1.27)		0.50 (1.10)		2.20 (1.49)	
.500	-2.28 (1.79)		-0.88 (1.35)		0.20 (0.90)		1.63 (1.70)		3.68 (0.98)
.625		-0.73 (1.57)		0.08 (1.23)		1.23 (1.21)		2.70 (1.38)	
.750			-0.05 (1.52)		0.43 (1.28)		2.08 (1.44)		
.875				0.33 (1.54)		1.68 (1.47)			
1.000					0.20 (1.40)				
Table									
.125				-2.03 (1.42)		-0.25 (1.76)			
.250			-1.90 (1.76)		-0.38 (1.35)		0.68 (1.79)		
.375		-2.20 (1.44)		-1.20 (1.31)		0.53 (1.40)		1.93 (1.99)	
.500	-3.00 (1.67)		-1.73 (1.28)		-0.03 (0.47)		1.65 (1.35)		3.10 (1.77)
.625		-1.83 (1.72)		-0.70 (1.10)		1.20 (1.03)		2.78 (0.88)	
.750			-0.53 (1.76)		0.58 (0.92)		1.98 (1.35)		
.875				-0.43 (1.56)		1.58 (1.20)			
1.000					0.35 (1.26)				

Figure Captions

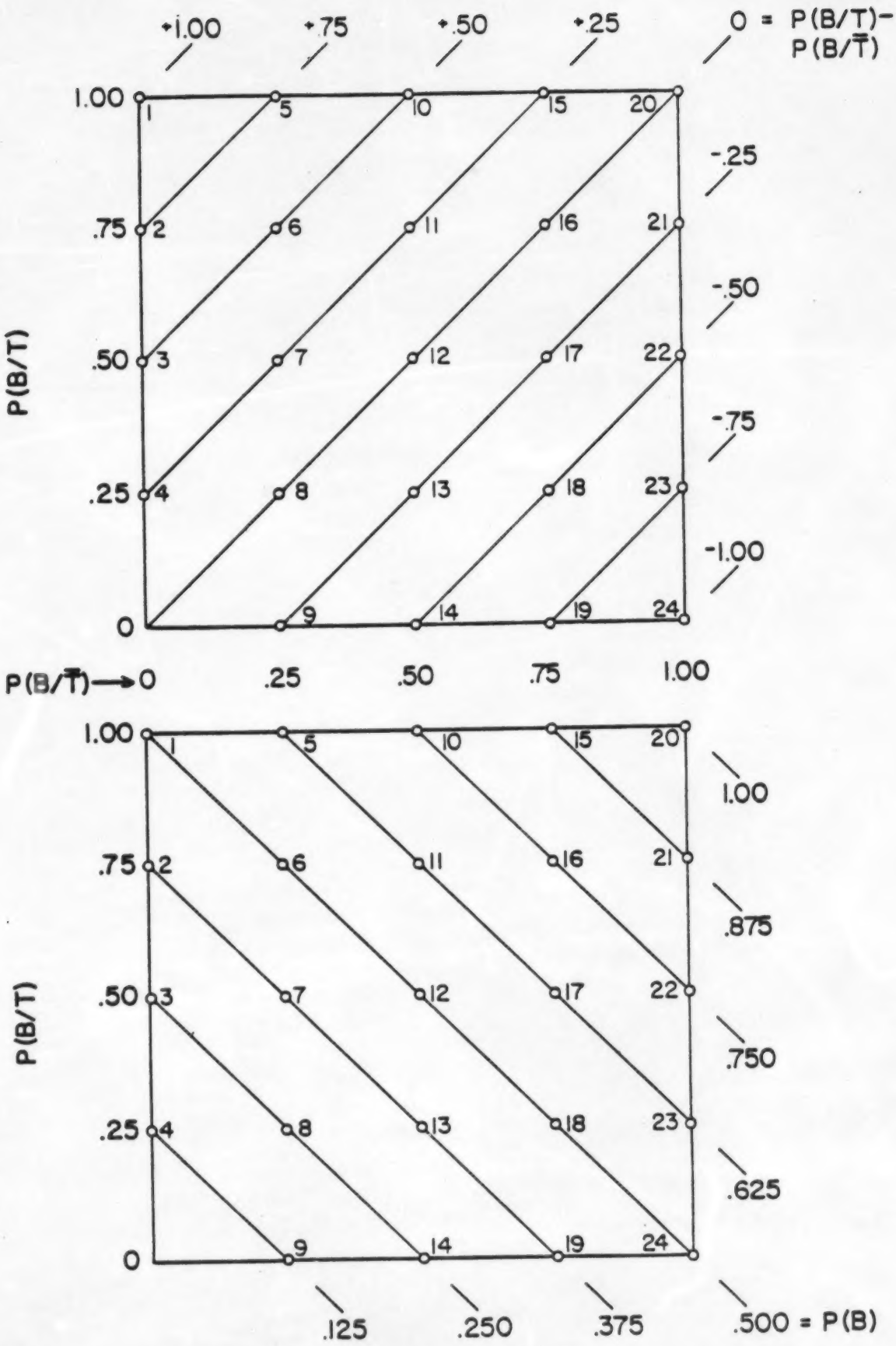
Figure 1. The 24 different response-outcome problems on the coordinates $p(B/T)$ and $p(B/\overline{T})$. The top portion locates the nine different response-outcome contingencies, $p(B/T) - p(B/\overline{T})$, on the unit square; the bottom portion locates the eight different outcome probabilities, $p(B)$. See text for additional explanation.

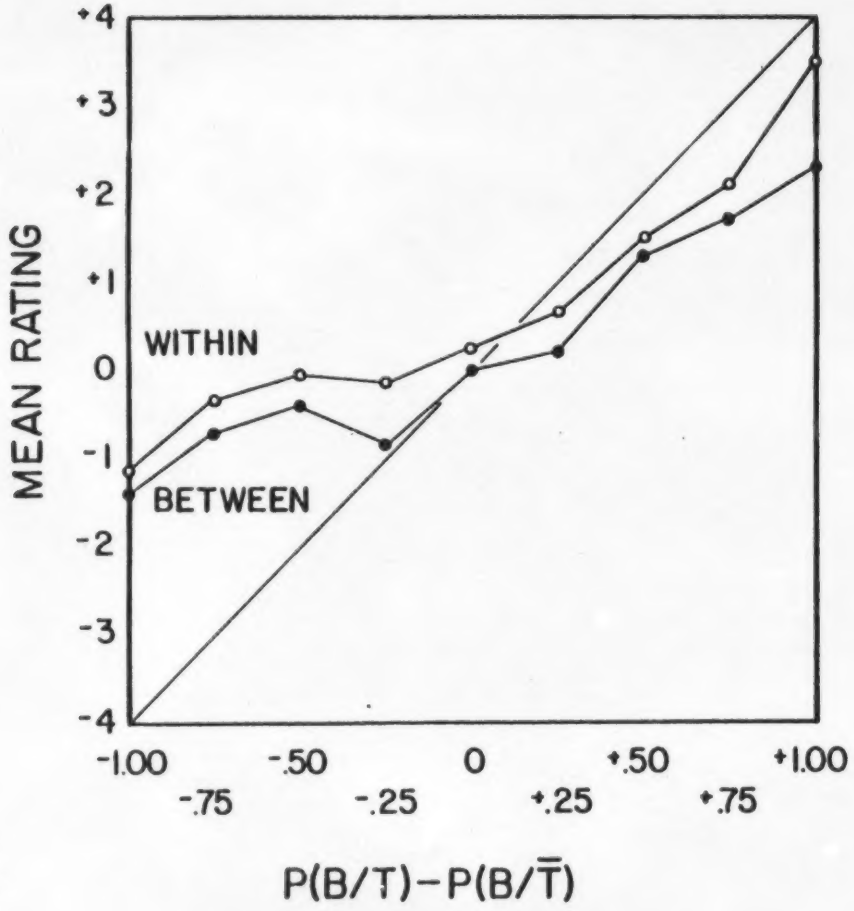
Figure 2. Contingency-judgment functions (left) and probability-judgment functions (right) in the within- and between-subjects parts of Experiment 1.

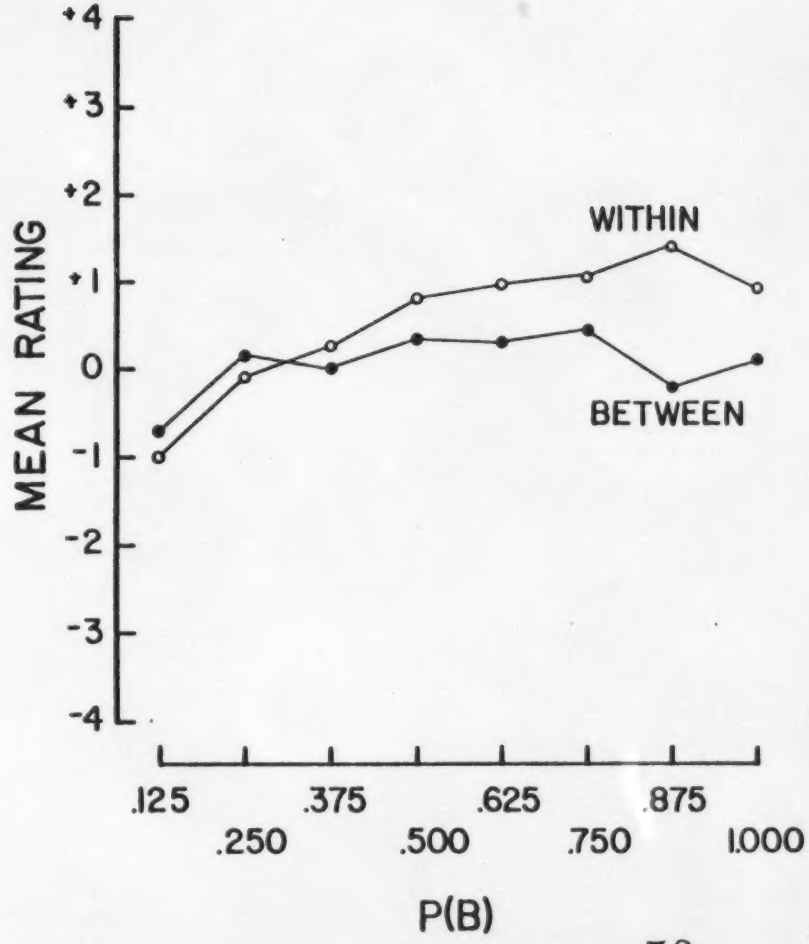
Figure 3. Contingency-judgment functions (left) and probability-judgment functions (right) under the time line and summary table formats of Experiment 2.

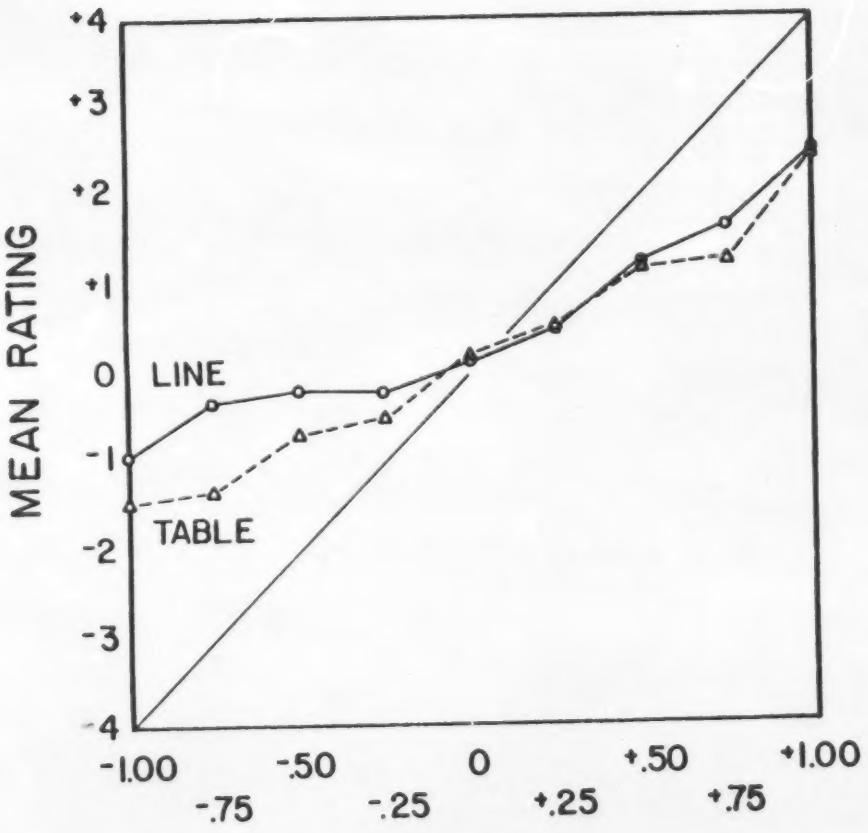
Figure 4. Contingency-judgment functions (left) and probability-judgment functions (right) under the broken and unbroken time line conditions of Experiment 3.

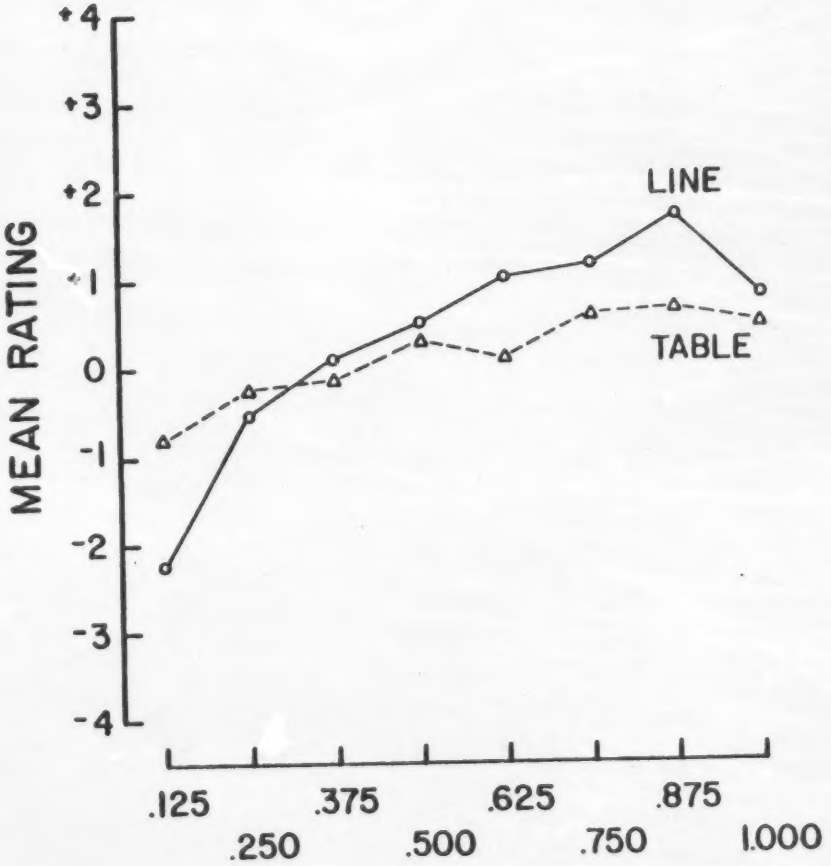
Figure 5. Contingency-judgment functions (left) and probability-judgment functions (right) under the four experimental conditions of Experiment 4.

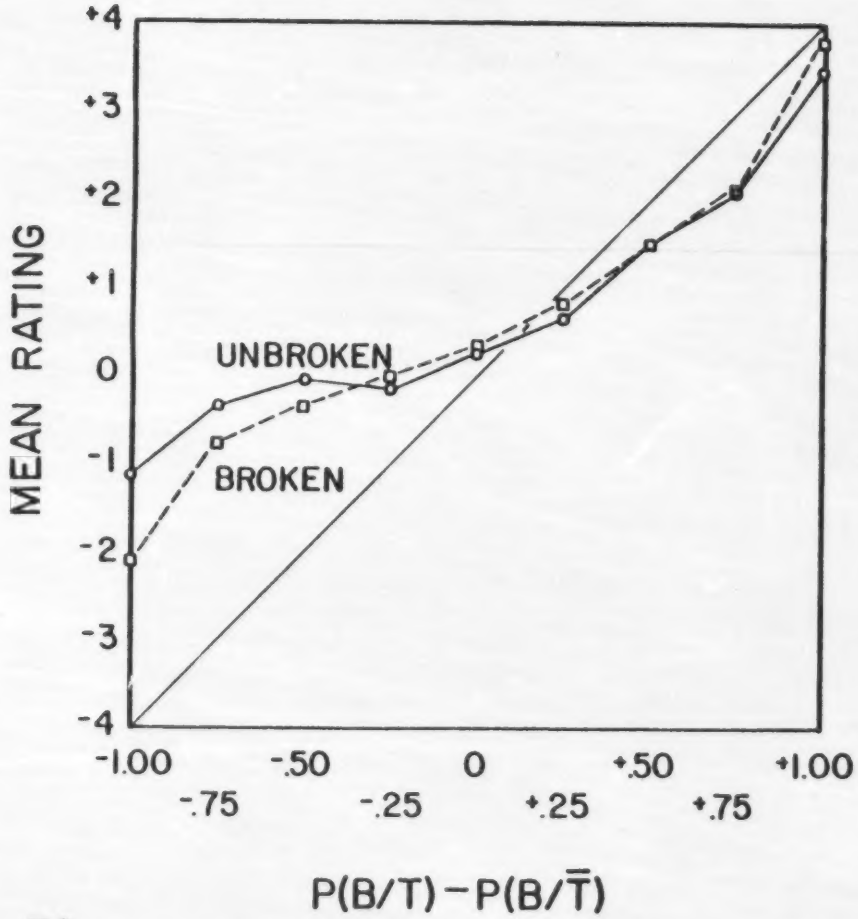

+1.00
+.75
+.50
+.25
0 = P(B/T)- P(B/T̄)
1.00
.75
.50
.25
0
P(B/T)
-.25
-.50
-.75
-1.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
P(B/T̄) → 0 .25 .50 .75 1.00
1.00
.875
.750
.625
.125
.250
.375
.500 = P(B)


MEAN RATING
+4
+3
+2
+1
0
-1
-2
-3
-4
WITHIN
BETWEEN
-1.00
-.75
-.50
-.25
0
+.25
+.50
+.75
+1.00
P(B/T)−P(B/T̄)

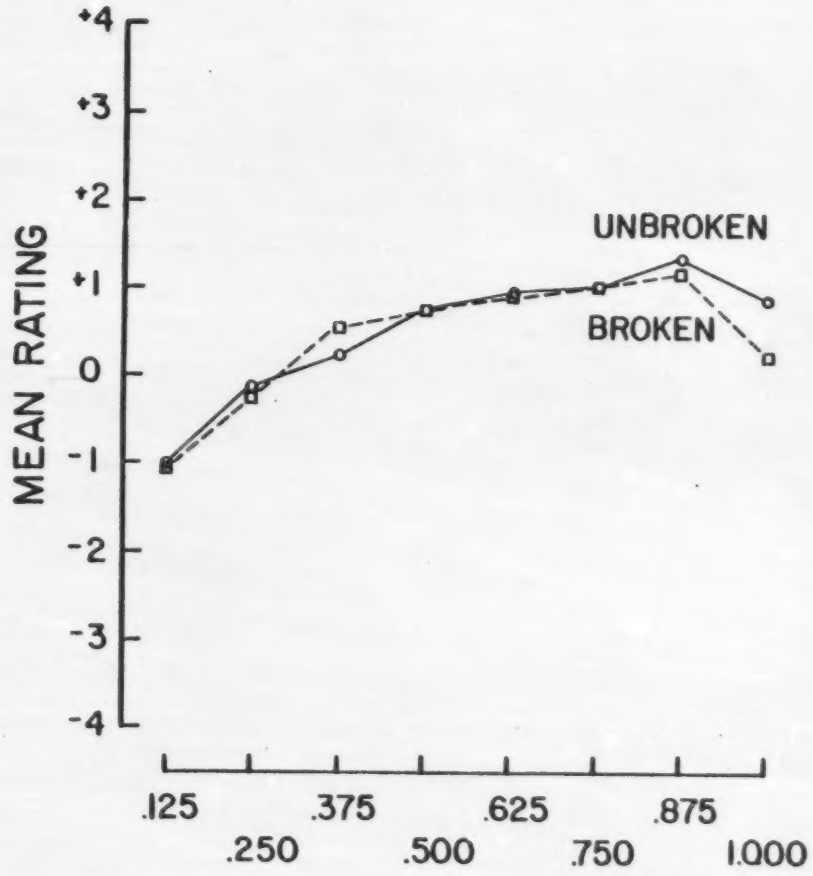

MEAN RATING
+4
+3
+2
+1
0
-1
-2
-3
-4
WITHIN
BETWEEN
.125
.250
.375
.500
.625
.750
.875
1.000
P(B)


MEAN RATING
+4
+3
+2
+1
0
-1
-2
-3
-4
LINE
TABLE
-1.00
-.75
-.50
-.25
0
+.25
+.50
+.75
+1.00
P(B/T)-P(B/$\bar{T}$)

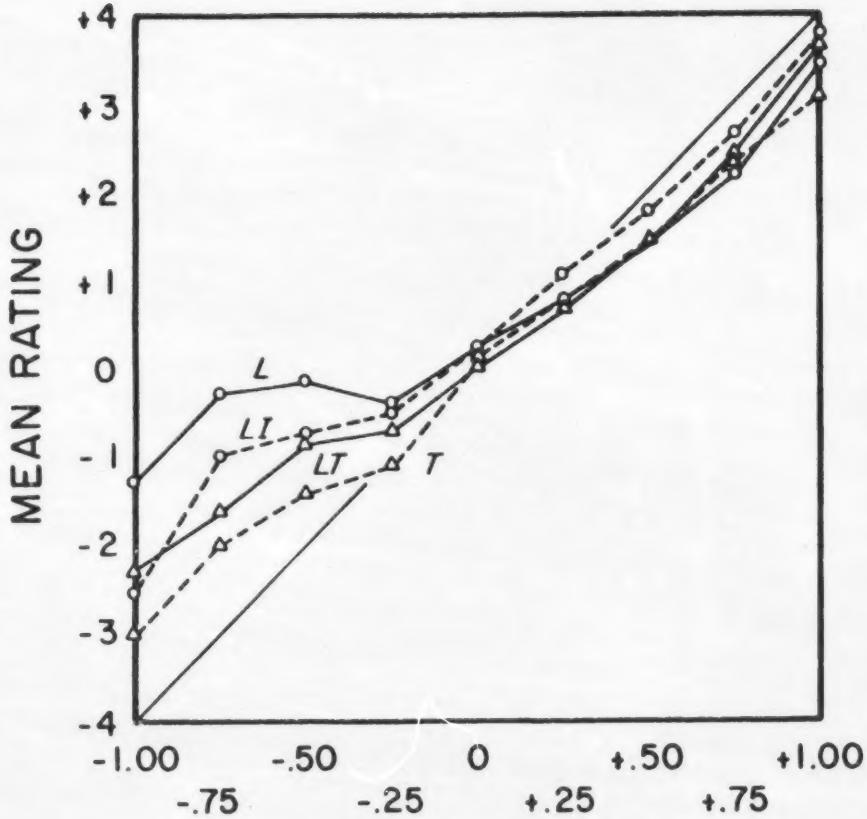

MEAN RATING
+4
+3
+2
+1
0
-1
-2
-3
-4
LINE
TABLE
.125
.250
.375
.500
.625
.750
.875
1.000
P(B)


MEAN RATING
+4
+3
+2
+1
0
-1
-2
-3
-4
UNBROKEN
BROKEN
-1.00
-.75
-.50
-.25
0
+.25
+.50
+.75
+1.00
P(B/T) − P(B/T̄)

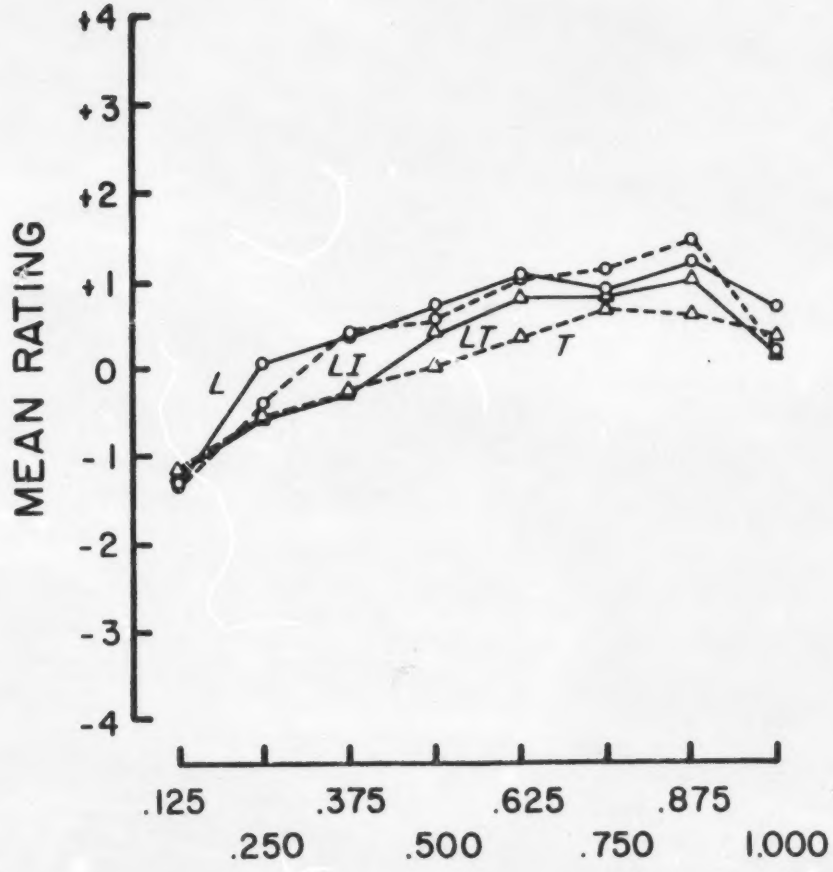
56


MEAN RATING
+4
+3
+2
+1
0
-1
-2
-3
-4
UNBROKEN
BROKEN
.125
.250
.375
.500
.625
.750
.875
1.000
P(B)

57


MEAN RATING
+4
+3
+2
+1
0
-1
-2
-3
-4
L
LI
LT
T
-1.00
-.75
-.50
-.25
0
+.25
+.50
+.75
+1.00
P(B/T) - P(B/T̄)

58


MEAN RATING
+4
+3
+2
+1
0
-1
-2
-3
-4
L
LI
LT
T
.125
.250
.375
.500
.625
.750
.875
1.000
P(B)

59